Bioinformatics

CS300

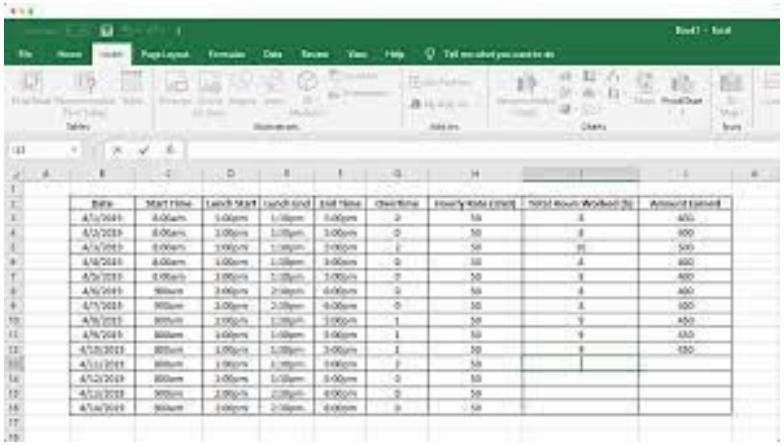
Building a Sequence Database

Week7, Deck 1

Fall 2022

Oliver BONHAM-CARTER

How to Store Data?

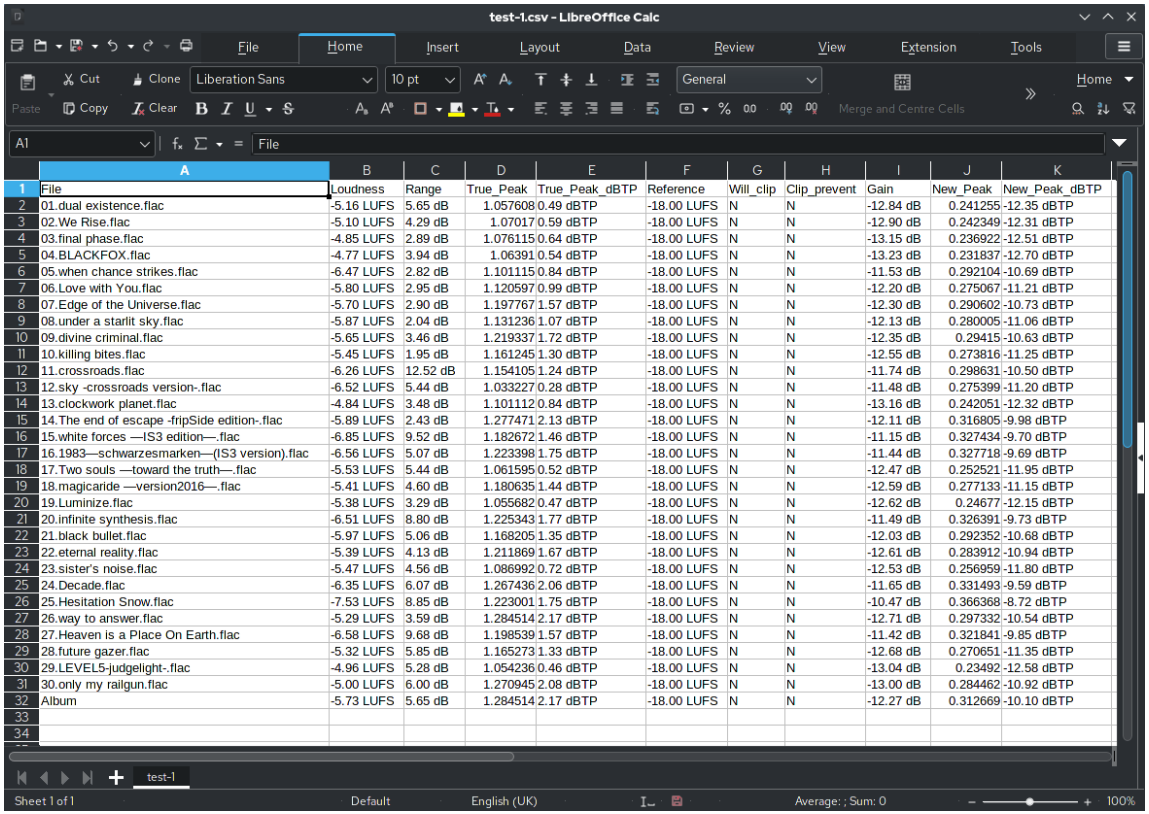


Top Tech Companies that Rule the World

	Symbol	Company	Entry Price	Last Price	% Change ↓	Market Cap (B)
1	NVDA	NVIDIA Corporation	$154.26	$191.17	19.31%	$116.33
2	AMD	Advanced Micro Devices	$23.53	$28.46	17.32%	$30.78
3	AAPL	Apple	$173.15	$207.16	16.42%	$976.82
4	AMZN	Amazon.com	$1,639.83	$1,901.75	13.77%	$935.76
5	F	Facebook	$161.45	$182.58	11.57%	$521.38
6	GOOG	Alphabet	$1,119.92	$1,256.00	10.83%	$874.70
7	MSFT	Microsoft Corporation	$112.03	$125.01	10.38%	$959.10
8	MTCH	Match Group	$55.38	$61.65	10.17%	$17.17
9	ATVI	Activision Blizzard	$42.14	$46.75	9.86%	$35.71
10	INTC	Intel Corporation	$52.96	$58.00	8.69%	$264.16
11	PYPL	PayPal Holdings	$98.07	$107.22	8.53%	$125.68
12	BABA	Alibaba Group	$183.03	$185.67	1.42%	$481.29
13	EA	Electronic Arts	$95.78	$94.00	-1.89%	$28.21
14	CR	salesforce.com	$163.65	$159.56	-2.56%	$123.64
15	TSLA	Tesla	$319.88	$258.00	-23.98%	$44.93

Based on "Tech Stocks That Move The Market" by Yahoo

body Spreadsheet



	File	Loudness	Range	True_Peak	True_Peak_dBTP	Reference	Will_clip	Clip_prevent	Gain	New_Peak	New_Peak_dBTP
2	01.dual existence.flac	-5.16 LUFS	5.65 dB	1.057608	0.49 dBTP	-18.00 LUFS	N	N	-12.84 dB	0.241255	-12.35 dBTP
3	02.We Rise.flac	-5.10 LUFS	4.29 dB	1.07017	0.59 dBTP	-18.00 LUFS	N	N	-12.90 dB	0.242349	-12.31 dBTP
4	03.final phase.flac	-4.85 LUFS	2.89 dB	1.076115	0.64 dBTP	-18.00 LUFS	N	N	-13.15 dB	0.236922	-12.51 dBTP
5	04.BLACKFOX.flac	-4.77 LUFS	3.94 dB	1.06391	0.54 dBTP	-18.00 LUFS	N	N	-13.23 dB	0.231837	-12.70 dBTP
6	05.when chance strikes.flac	-6.47 LUFS	2.82 dB	1.101115	0.84 dBTP	-18.00 LUFS	N	N	-11.53 dB	0.292104	-10.69 dBTP
7	06.Love with You.flac	-5.80 LUFS	2.95 dB	1.120597	0.99 dBTP	-18.00 LUFS	N	N	-12.20 dB	0.275067	-11.21 dBTP
8	07.Edge of the Universe.flac	-5.70 LUFS	2.90 dB	1.197767	1.57 dBTP	-18.00 LUFS	N	N	-12.30 dB	0.290602	-10.73 dBTP
9	08.under a starlit sky.flac	-5.87 LUFS	2.04 dB	1.131236	1.07 dBTP	-18.00 LUFS	N	N	-12.13 dB	0.280005	-11.06 dBTP
10	09.divine criminal.flac	-5.65 LUFS	3.46 dB	1.219337	1.72 dBTP	-18.00 LUFS	N	N	-12.35 dB	0.29415	-10.63 dBTP
11	10.killing bites.flac	-5.45 LUFS	1.95 dB	1.161245	1.30 dBTP	-18.00 LUFS	N	N	-12.55 dB	0.273816	-11.25 dBTP
12	11.crossroads.flac	-6.26 LUFS	12.52 dB	1.154105	1.24 dBTP	-18.00 LUFS	N	N	-11.74 dB	0.298631	-10.50 dBTP
13	12.sky -crossroads version-.flac	-6.52 LUFS	5.44 dB	1.033227	0.28 dBTP	-18.00 LUFS	N	N	-11.48 dB	0.275399	-11.20 dBTP
14	13.clockwork planet.flac	-4.84 LUFS	3.48 dB	1.101112	0.84 dBTP	-18.00 LUFS	N	N	-13.16 dB	0.242051	-12.32 dBTP
15	14.The end of escape -fripSide edition-.flac	-5.89 LUFS	2.43 dB	1.277471	2.13 dBTP	-18.00 LUFS	N	N	-12.11 dB	0.316805	-9.98 dBTP
16	15.white forces —IS3 edition—.flac	-6.85 LUFS	9.52 dB	1.182672	1.46 dBTP	-18.00 LUFS	N	N	-11.15 dB	0.327434	-9.70 dBTP
17	16.1983—schwarzesmarken—(IS3 version).flac	-6.56 LUFS	5.07 dB	1.223398	1.75 dBTP	-18.00 LUFS	N	N	-11.44 dB	0.327718	-9.69 dBTP
18	17.Two souls —toward the truth—.flac	-5.53 LUFS	5.44 dB	1.061595	0.52 dBTP	-18.00 LUFS	N	N	-12.47 dB	0.252521	-11.95 dBTP
19	18.magicaride —version2016—.flac	-5.41 LUFS	4.60 dB	1.180635	1.44 dBTP	-18.00 LUFS	N	N	-12.59 dB	0.277133	-11.15 dBTP
20	19.Luminize.flac	-5.38 LUFS	3.29 dB	1.055682	0.47 dBTP	-18.00 LUFS	N	N	-12.62 dB	0.24677	-12.15 dBTP
21	20.infinite synthesis.flac	-6.51 LUFS	8.80 dB	1.225343	1.77 dBTP	-18.00 LUFS	N	N	-11.49 dB	0.326391	-9.73 dBTP
22	21.black bullet.flac	-5.97 LUFS	5.06 dB	1.168205	1.35 dBTP	-18.00 LUFS	N	N	-12.03 dB	0.292352	-10.68 dBTP
23	22.eternal reality.flac	-5.39 LUFS	4.13 dB	1.211869	1.67 dBTP	-18.00 LUFS	N	N	-12.61 dB	0.283912	-10.94 dBTP
24	23.sister's noise.flac	-5.47 LUFS	4.56 dB	1.086992	0.72 dBTP	-18.00 LUFS	N	N	-12.53 dB	0.256959	-11.80 dBTP
25	24.Decade.flac	-6.35 LUFS	6.07 dB	1.267436	2.06 dBTP	-18.00 LUFS	N	N	-11.65 dB	0.331493	-9.59 dBTP
26	25.Hesitation Snow.flac	-7.53 LUFS	8.85 dB	1.223001	1.75 dBTP	-18.00 LUFS	N	N	-10.47 dB	0.366368	-8.72 dBTP
27	26.way to answer.flac	-5.29 LUFS	3.59 dB	1.284514	2.17 dBTP	-18.00 LUFS	N	N	-12.71 dB	0.297332	-10.54 dBTP
28	27.Heaven is a Place On Earth.flac	-6.58 LUFS	9.68 dB	1.198539	1.57 dBTP	-18.00 LUFS	N	N	-11.42 dB	0.321841	-9.85 dBTP
29	28.future gazer.flac	-5.32 LUFS	5.85 dB	1.165273	1.33 dBTP	-18.00 LUFS	N	N	-12.68 dB	0.270651	-11.35 dBTP
30	29.LEVEL5-judgelight-.flac	-4.96 LUFS	5.28 dB	1.054236	0.46 dBTP	-18.00 LUFS	N	N	-13.04 dB	0.23492	-12.58 dBTP
31	30.only my railgun.flac	-5.00 LUFS	6.00 dB	1.270945	2.08 dBTP	-18.00 LUFS	N	N	-13.00 dB	0.284462	-10.92 dBTP
32	Album	-5.73 LUFS	5.65 dB	1.284514	2.17 dBTP	-18.00 LUFS	N	N	-12.27 dB	0.312669	-10.10 dBTP

Spreadsheets?

Good for Small Data Sets

- **Pros**
  - Convenient to maintain a spreadsheet
  - Lots of data available online in spreadsheets
  - Looking up specifics

- **Cons**
  - Hard to query for specifics
  - Data appears crowded and confusing
  - File limits on how much data you can store.
  - Security: spreadsheet make all data available to all users?

So, Use a Database!!

- **Improved data sharing and data security**
- **Effective data integration**
- **Consistent, reliable data**
- **Maintainable**
- **Increased productivity**
- **Better decision-making**
- **Databases designed to maintain lots of data!**
- **And more!**

Let’s Make a Database!!

- Simple, easy to learn SQL language
- Open source
- Built for sophisticated queries to find *any* type of data.
- Download Link: https://www.sqlite.org/download.html

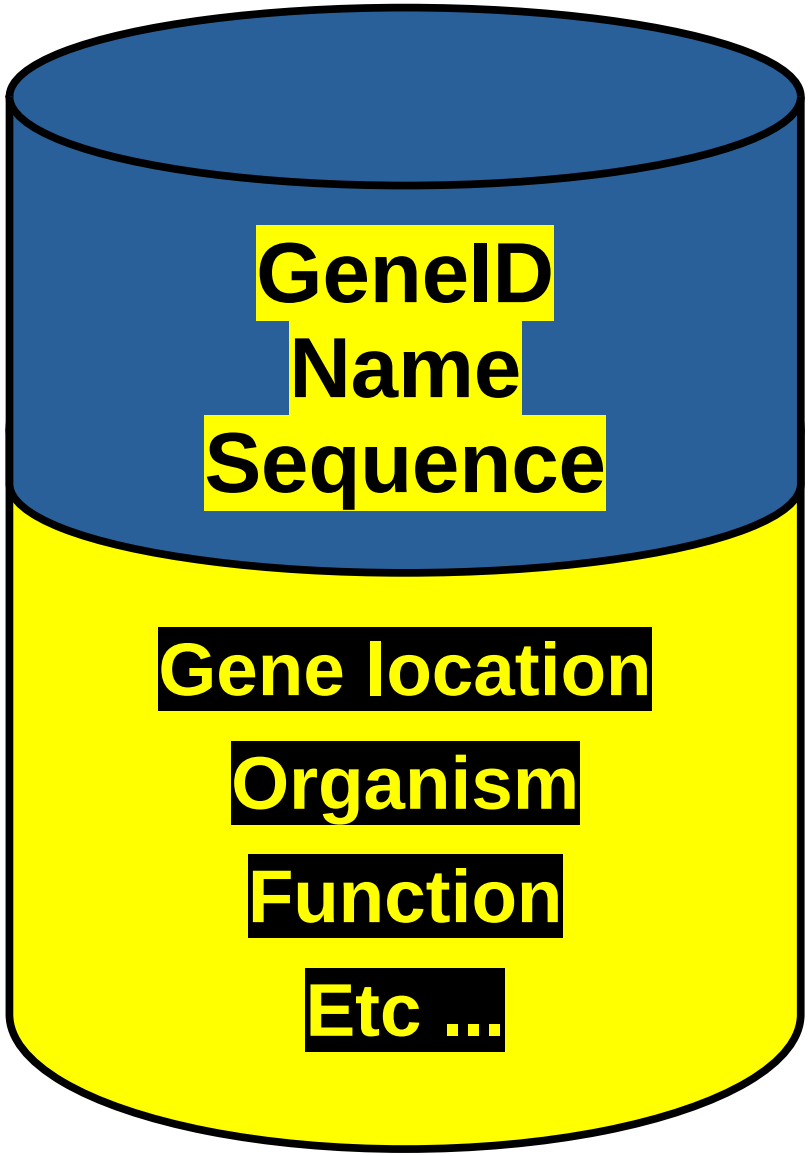
SQLite Resources

- **Documentation**: https://www.sqlite.org/docs.html
- **Online SQLite3 (demonstrations)**
  - https://extendsclass.com/sqlite-browser.html
  - https://sqliteonline.com/
- **Command Line Reference**
  - https://www.sqlite.org/cli.html

Coding Time!

Want to Store … !

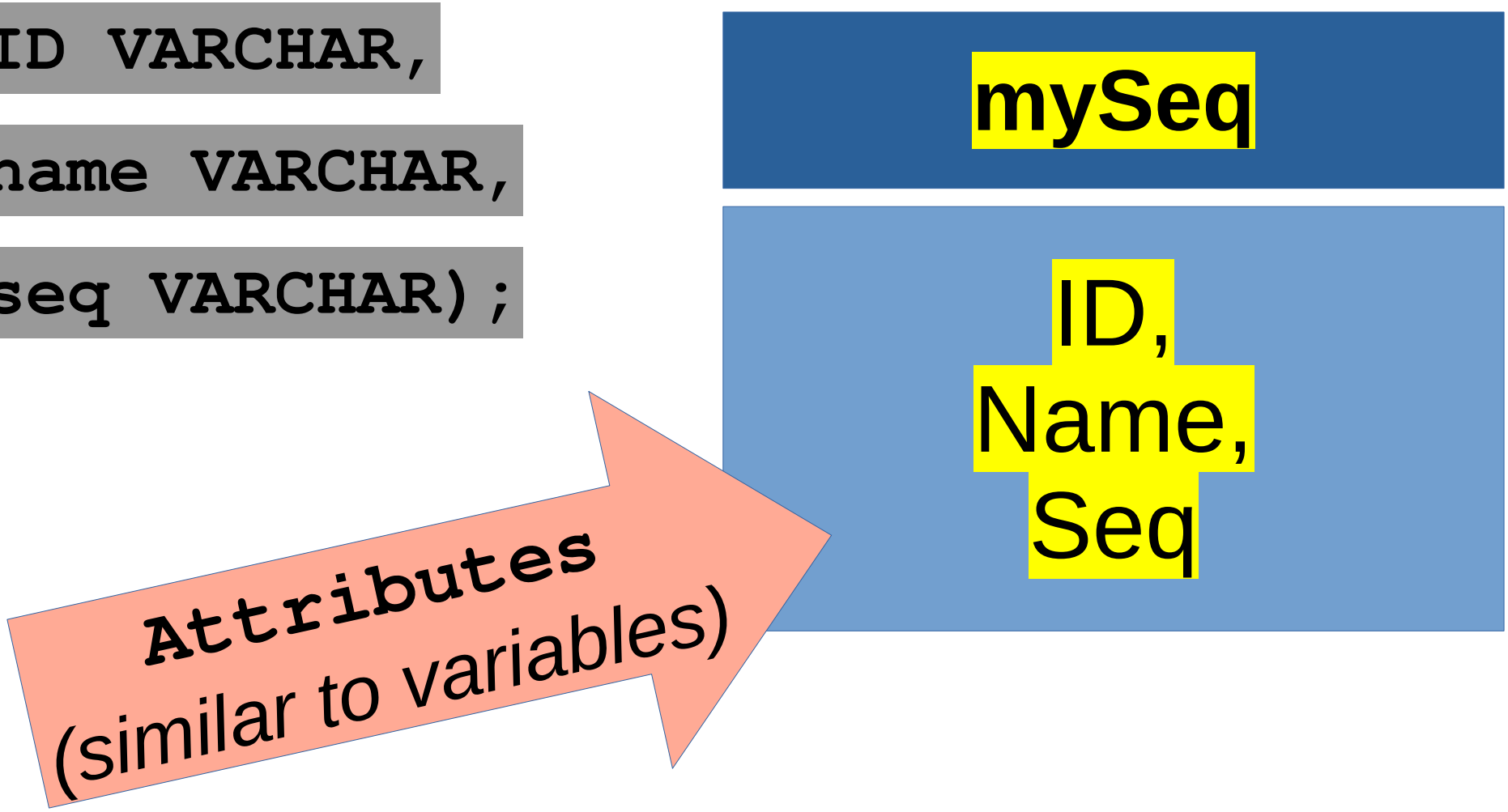
- Store this information
- Maybe this information too?



Basic Schema

For a single table

```
CREATE TABLE mySeq(
ID VARCHAR,
name VARCHAR,
seq VARCHAR);
```



Create a Table

- Paste in your table creation code.
- Click “Run”

SQLite Browser Online:
https://extendsclass.com/sqlite-browser.html

Add Data (*Populate*)

```
INSERT INTO mySeq VALUES
(
"gene101",
"x-gene",
"ATATCG"
);
```

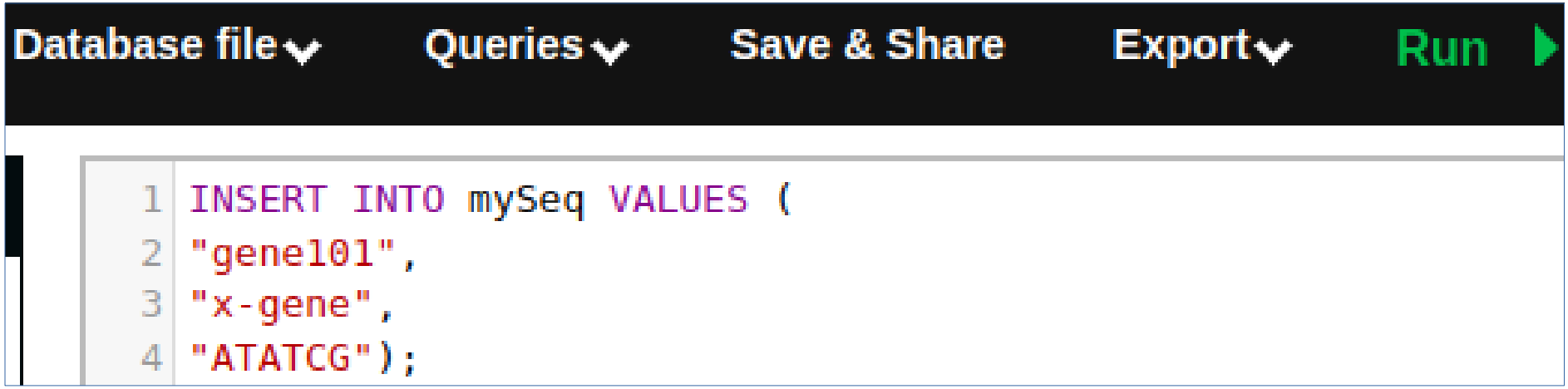
mySeq

ID = “gene101”,
Name = “x-gene”,
Seq = “ATATCG”

Stored as a tuple

Add Data (*Populate*)

- Paste insert code with data as a tuple
- Click “Run”



Query **mySeq** Table

```
SELECT * FROM mySeq;
```

mySeq

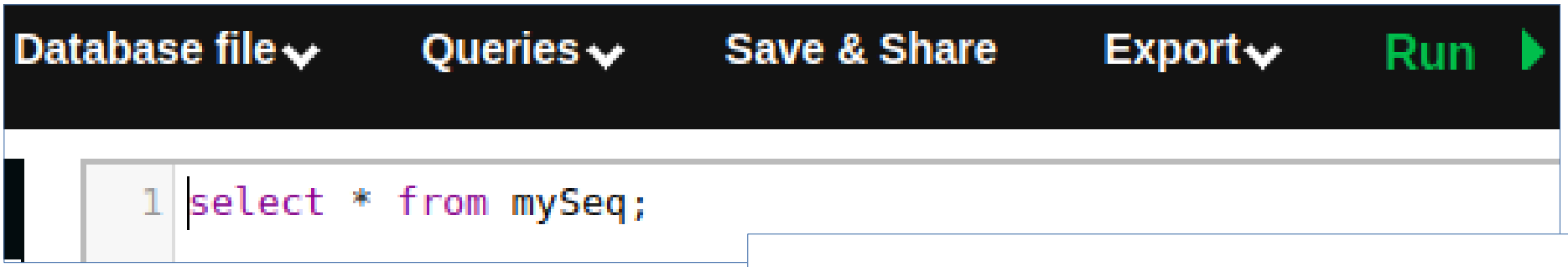
ID = “gene101”,
Name = “x-gene”,
Seq = “ATATCG”

Query

ID	name	seq
gene101	x-gene	ATATCG

Query **`mySeq`** Table

- Paste query code
- Click “Run”

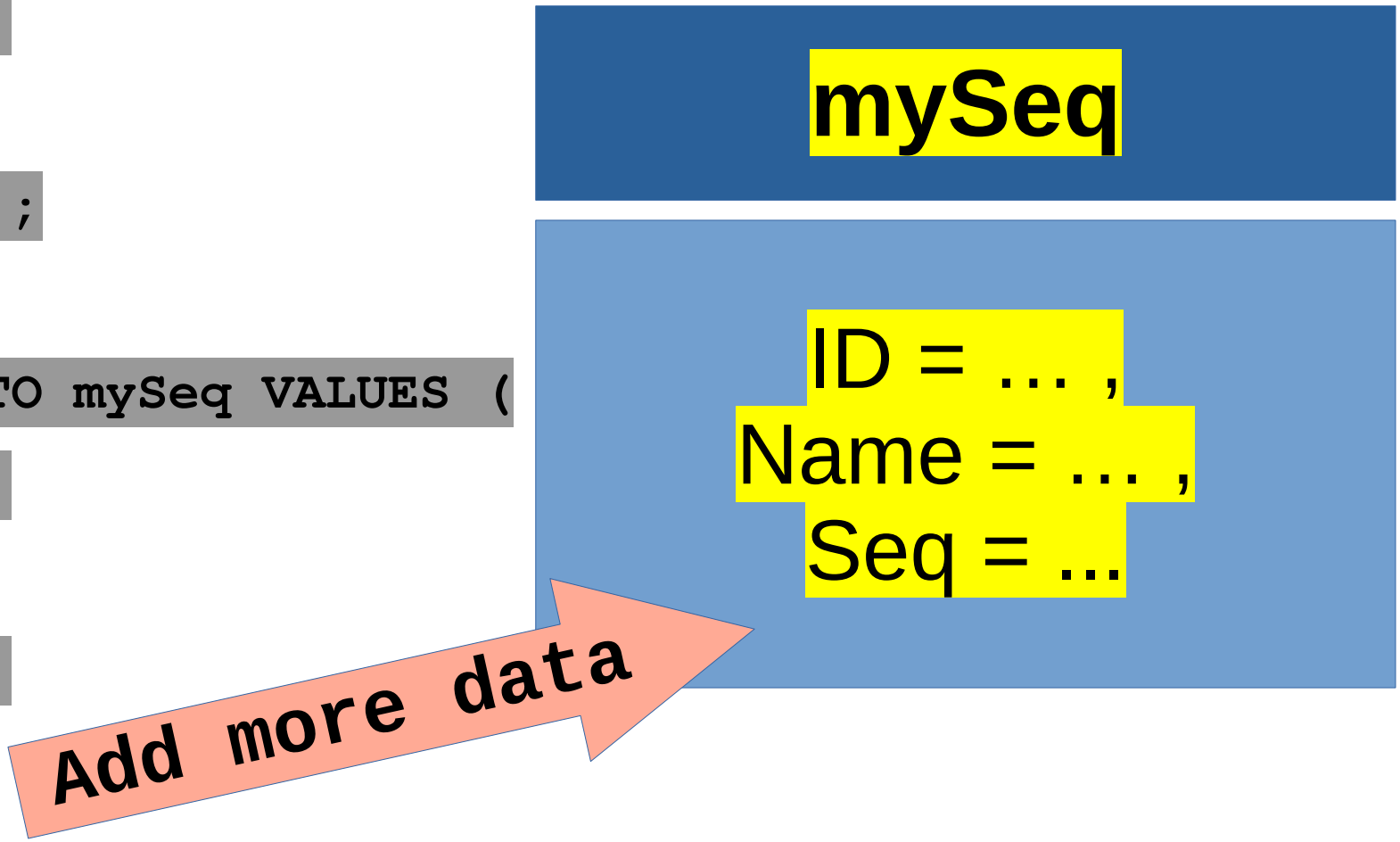


ID	name	seq
gene101	x-gene	ATATCG

Add More Data

```
INSERT INTO mySeq VALUES (
"gene210",
"a-gene",
"GATATCG");
```

```
INSERT INTO mySeq VALUES (
"gene300",
"b-gene",
"GTATCG");
```



Query ID's and Seqs

```
SELECT ID,seq FROM mySeq;
```

mySeq

ID = "gene101",
Name = "x-gene",
Seq = "ATATCG"

Query

ID	seq
gene101	ATATCG
gene210	GATATCG
gene300	GTATCG

Building Your Own Database!

- Create a database to contain the following information.
  - `ID`
  - `Organism`
  - `DNASeq`
  - `Protseq`
- Add at least four rows of data (this data can be real or made-up)
- Offer two queries for two different types of information from the data.

Building Your Own Database!

GitHub Classroom working repository:
https://classroom.github.com/a/pqKAOxgT

Please Push

- **Table creation code**
- **Code to insert data**
- **SQL query code for two queries**
- **Sample Output**

Due on Friday

14 Oct 2022